

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-056007

(43)Date of publication of application : 05.03.1993

(51)Int.Cl. H04B 14/04  
G10L 3/00

(21)Application number : 03-211807

(71)Applicant : NIPPON HOSO KYOKAI <NHK>

(22)Date of filing : 23.08.1991

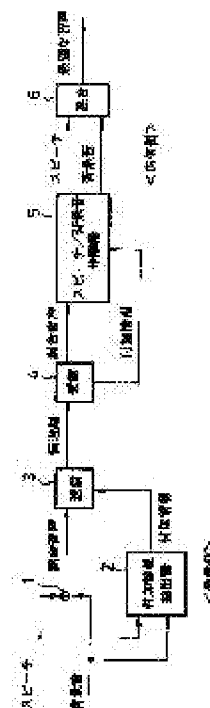
(72)Inventor : IMAI TORU  
MIYASAKA EIICHI  
UMEDA TETSUO

## (54) MIXED VOICE SIGNAL TRANSMISSION SYSTEM

### (57)Abstract:

**PURPOSE:** To separate a speech signal and a background sound from a mixed voice signal with high sound quality and to mix the both in a preferred rate by the listener while suppressing a capacity required for additional information.

**CONSTITUTION:** A sender side mixes a speech signal and a background sound, generates additional information for separating signals (by additional information extraction section 2) from the speech signal and the background sound and multiplexes the mixed voice signal and the additional information signal with each other and sends the multiplexed signal. A receiver side separates the mixed voice signal and the additional information from the received signal, separates the speech signal and the background sound based on the additional information (speech/background separator 5), and mixes the separated speech signal and background sound in a rate preferred by the listener.



(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平5-56007

(43)公開日 平成5年(1993)3月5日

(51)Int.Cl. <sup>5</sup>	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 B 14/04	Z	4101-5K		
G 1 0 L 3/00	N	8946-5H		

審査請求 未請求 請求項の数 1 (全 6 頁)

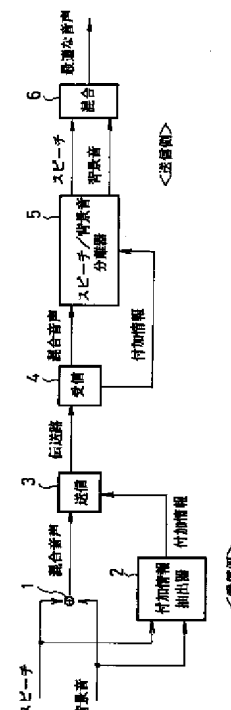
(21)出願番号	特願平3-211807	(71)出願人	000004352 日本放送協会 東京都渋谷区神南2丁目2番1号
(22)出願日	平成3年(1991)8月23日	(72)発明者	今井 亨 東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
		(72)発明者	宮坂 栄一 東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
		(72)発明者	梅田 哲夫 東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
		(74)代理人	弁理士 三好 秀和 (外5名)

(54)【発明の名称】 混合音声信号伝送方式

(57)【要約】

【目的】 付加情報の容量を少なく抑えつつ、混合音声からスピーチと背景音とを高音質で分離でき、聴取者が自分の好みの割合で両者を混合することを可能とする。

【構成】 送信側では、スピーチと背景音とを混合するとともに、スピーチと背景音とから信号分離用の付加情報を生成(付加情報抽出器2)してこれら混合音声信号と付加情報信号とを多重して伝送する一方、受信側では、受信信号から混合音声と付加情報とを分離するとともに、付加情報に基づいてスピーチと背景音を分離(スピーチ/背景音分離器5)し、分離したスピーチと背景音は聴取者の好みの割合で混合される。



## 【特許請求の範囲】

【請求項1】 送信側では、スピーチと背景音とを混合するとともに、スピーチと背景音とから信号分離用の付加情報を生成してこれら混合音声信号と付加情報信号とを多重して伝送する一方、受信側では、受信信号から混合音声と付加情報とを分離するとともに、付加情報に基づいてスピーチと背景音を分離し、分離したスピーチと背景音は聴取者の好みの割合で混合されるようにしたことを特徴とする混合音声信号伝送方式。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、聴覚障害者や高齢者等の音声聴取に好適な混合音声信号伝送方式に関する。

【0002】〔発明の概要〕本発明は、スピーチ（人の声）と、音楽や効果音等の背景音とが混在した音声信号を送信する放送システム等において、スピーチと背景音に関する時間情報やスペクトル情報等の付加情報を少ない量子化ビット数により多重して伝送し、受信側では、付加情報に基づいてフィルタ等の処理を混合音声に施すことにより、混合音声からスピーチと背景音とを分離し、受信者が自分の好みの混合比により両者を再び混合して聴取できるようにしたものである。

## 【0003】

【従来の技術】音声に混在する雑音除去あるいは特定の音声のみを抽出することを目的とする従来技術としては、雑音源の情報を利用する2入力方式と、混在された音声しか利用できない1入力方式とが知られている。

【0004】2入力方式としては、適応フィルタを用いた雑音除去方法が知られている（例えば、B.Widrow et al.: "Adaptive Noise Cancelling: Principles and Applications", Proc. IEEE, Vol. 63, No. 12, pp. 1692-1716 (1975)）。

【0005】1入力方式としては、音声信号が存在しない区間を何らかの方法により検出し、その区間から雑音を推定してウィナーフィルタ（例えば、N. Wiener:

"Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications", John Wiley & Sons, (1949)）やスペクトルサブトラクション法（S.F. Boll: "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans., Vol. ASSP-27, No. 2, April (1979)）により雑音除去を行う。

## 【0006】

【発明が解決しようとする課題】ところで、聴覚障害者や高齢者には、放送の番組音声、特に人の声が聞き取りにくいことが多い。特に、番組効果のために音楽や効果音がスピーチと混合された場合、極めて聞き取りにくくなる。

【0007】これを解決するために、上記従来の1入力方式を適用して、受信した音声だけからスピーチを分離

することも考えられるが、放送システムでは様々な音声を扱うため、非常に困難である。また、2入力方式により、スピーチや背景音の情報を送ることも考えられるが、放送等のシステムではその伝送容量に限りがある。例えば、通常の音声信号の他にスピーチの音声波形そのものを伝送しようとする、当然さらに1チャンネル分の伝送容量が必要となり、効率的な伝送を行うには好ましくない。

【0008】従って、現行の放送システム、特に音声の放送における受信者サービスには、ステレオ放送、2か国語放送、音声多重放送等、音声チャンネルを増やすことによる付加価値的な音声のサービスはあるものの、聴覚障害者や高齢者など、通常の番組音声聞き辛い受信者を対象に、情報伝達の本質であるスピーチを聞きやすくするための配慮はなされていないのが現状である。

【0009】本発明は、上記事情に鑑みてなされたものであり、その目的は、付加情報の容量を少なく抑えたまま高音質でスピーチと背景音とを分離できるとともに、聴取者が自分の好みの割合で両者を混合することのできる混合音声信号伝送方式を提供することにある。

## 【0010】

【課題を解決するための手段】上記の目的を達成するために本発明に係る混合音声信号伝送方式は、送信側では、スピーチと背景音とを混合するとともに、スピーチと背景音とから信号分離用の付加情報を生成してこれら混合音声信号と付加情報信号とを多重して伝送する一方、受信側では、受信信号から混合音声と付加情報とを分離するとともに、付加情報に基づいてスピーチと背景音を分離し、分離したスピーチと背景音は聴取者の好みの割合で混合されるようにしたことを特徴とする。

## 【0011】

【作用】送信側において、スピーチの各フレームが、無音区間か、周期的区間か、あるいは非周期的区間かという区間情報、スピーチが周期的区間であるときのピッチデータ、スピーチが有音区間であるときのスピーチのパワースペクトル包絡を表すベクトル番号、およびスピーチが有音区間であるときの背景音のパワースペクトル包絡を表すベクトル番号等の付加情報が生成され、混合音声信号とともに多重伝送される。

【0012】受信側では、まず、混合音声と付加情報とが分離され、この付加情報に基づき、混合音声からスピーチと背景音に分離される。分離されたスピーチと背景音は聴取者の好みの割合で混合された音声となる。

## 【0013】

【実施例】図1は、本発明方式が適用された一実施例構成を示すブロック図である。

【0014】図示の装置の送信系は、音声混合器1と、付加情報抽出器2と、送信機3とを備え、スピーチと背景音とを混合するとともに、スピーチと背景音とから付加情報を生成してこれら混合音声信号と付加情報信号と

3

を多重して伝送する。一方の受信系は、受信機4と、スピーチ／背景音分離器5と、音声混合器6とを備え、受信信号から混合音声と付加情報とを分離するとともに、付加情報に基づいてスピーチと背景音を分離し、分離したスピーチと背景音は好みの割合で混合される。

【0015】図2には、上記付加情報抽出器2の構成が示されている。

【0016】この付加情報抽出器2は、スピーチの信号からフレームを切り出すフレーム切り出し部201と、スピーチ信号のフレームから無音・有音を判定する無音・有音判定部202と、有音区間のピッチを抽出するピッチ抽出部203と、同じく有音区間からフレームの線形予測分析を行うLPC分析部204と、LPC分析部204のスペクトル包絡の係数をベクトル量子化するベクトル量子化部205と、背景音の信号からフレームを切り出すフレーム切り出し部206と、切り出されたフレームの線形予測分析を行うLPC分析部207と、LPC分析部207のスペクトル包絡の係数をベクトル量子化するベクトル量子化部208と、無音・有音判定部202の区間情報とピッチ抽出部203のピッチデータと各ベクトル量子化部205および208の各ベクトル番号とから符号化された付加情報を生成して前記送信機3へ出力するエンコーダ209とを備えている。

【0017】次に、この付加情報抽出器2の作用を説明する。

【0018】スピーチは、フレーム切り出し部201において、フレーム長20ms、フレーム周期10msのハミング窓により一定長のフレームに切り出される。切り出された各フレームは無音・有音判定部202に出力される。

【0019】無音・有音判定部202では、供給された各フレームが無音区間か有音区間かの判定をそのパワーに基づいて行う。ここでいう無音区間とは、スピーチの無音区間、すなわち言葉を話していない区間のことをいい、有音区間とは無音区間以外の区間をいう。この区間情報はエンコーダ209に出力される。

【0020】有音区間については、ピッチ抽出部203においてそのピッチが抽出される。ピッチが抽出された場合はこれを周期的区間とし、ピッチが抽出されない場合には非周期的区間とする。周期的区間のピッチデータはエンコーダ209に出力される。

【0021】また、有音区間と判定されたフレームについては、LPC分析部204において線形予測分析がされる。ここでは、5次のLPCケプストラム係数を用いパワースペクトルの包絡を表す係数が求められる。こうして求められたスペクトル包絡の係数はベクトル量子化部205に出力される。

【0022】ベクトル量子化部205では、スペクトル包絡の係数がベクトル量子化されベクトル番号に変換される。なお、ベクトル量子化のためのコードブックは、

4

予め多数のスピーチから作成しておく。また、コードブックのサイズは8ビット256とし、LBG (Linde, Buza, Gray) アルゴリズムを用いてクラスタリングを行う。このとき、真のスペクトルとコードブック中で最も近いスペクトルとの距離が、ある一定のしきい値よりも大きいときは量子化しない。求められたベクトル番号はエンコーダ209に出力される。

【0023】一方、背景音に関しても、スピーチ同様に、フレーム切り出し部206でフレーム切り出しが行われ、切り出された各フレームの線形予測分析がLPC分析部207で行われる。そのスペクトル包絡の係数はベクトル量子化部208でベクトル番号に変換される。ただし、ベクトル量子化のためのコードブックは、前記ベクトル量子化部205におけるコードブックとは異なり、予め音楽や効果音等の多数の背景音から作成される。求められたベクトル番号はエンコーダ209に出力される。

【0024】エンコーダ209には、スピーチの各フレームが、無音区間か、周期的区間か、あるいは非周期的区間かという区間情報、スピーチが周期的区間であるときのピッチデータ、スピーチが有音区間であるときのスピーチのパワースペクトル包絡を表すベクトル番号、およびスピーチが有音区間であるときの背景音のパワースペクトル包絡を表すベクトル番号が供給され、これらの情報は符号化された付加情報となる。

【0025】こうして求められた付加情報は前記送信機3において混合音声に多重されて伝送される。

【0026】図3には、前記スピーチ／背景音分離器5の構成が示されている。

【0027】このスピーチ／背景音分離器5は、入力した付加情報から前記区間情報、ピッチデータ、スピーチのベクトル番号および背景音のベクトル番号を復号化するデコーダ501と、復号化された区間情報から混合音声のフレーム切り出しを行うフレーム切り出し部502と、フレームの区間を判定する区間判定部503と、無音区間の混合音声を除く除去部504と、有音区間の混合音声を高速フーリエ変換するFFT505と、FFT出力から背景音のパワースペクトルを減算するスペクトルサブトラクション部506と、スペクトルサブトラクション部506の出力スペクトルをフィルタリングするウィナーフィルタ507と、ウィナーフィルタ出力を高速逆フーリエ変換するIFFT508と、IFFT出力をフィルタリングするくし形フィルタ509と、フレームのつなぎ合わせを行うつなぎ合わせ部510と、適応フィルタ部511とを備えて構成されている。

【0028】次にこのスピーチ／背景音分離器5の作用を説明する。

【0029】受信機4により受信信号が混合音声と付加情報に分離された後、スピーチ／背景音分離器5に供給

される。

【0030】デコード部501では、付加情報から前記区間情報、ピッチデータ、スピーチのベクトル番号および背景音のベクトル番号がデコードされる。区間情報はフレーム切り出し部502と区間判定部503へ、ピッチデータはくし形フィルタ部509へ、スピーチのベクトル番号はウィナーフィルタ部507へ、背景音のベクトル番号はスペクトルサブトラクション部506へそれぞれ出力される。

【0031】フレーム切り出し部502では、入力した混合音声10がフレーム単位で切り出される。切り出しタイミングは区間情報を参照しながら送信側と同期をとって行う。

【0032】区間判定部503では、フレームに切り出された混合音声10が区間情報を参照してスピーチの無音区間、有音区間、周期的区間、非周期的区間に分離される。分離された無音区間は除去部504へ出力され、この除去部504にて無音区間の混合音声10が除去される。除去された無音区間の混合音声はつなぎ合わせ部510に出力される。

【0033】有音区間については、FFT部505の高速フーリエ変換によって周波数領域に変換され、混合音声のパワースペクトルが求められる。

【0034】スペクトルサブトラクション部506では、混合音声のパワースペクトルから背景音のベクトル番号に対応したパワースペクトルが減算され、スピーチのパワースペクトルが求められる。このとき、減算結果が負となるところは0に置き換える。

【0035】次に、ウィナーフィルタ部507では、入力したスピーチのベクトル番号に基づきスペクトルサブトラクション部506の出力に対して次式で示される特性を持つウィナーフィルタ処理を施す。

【0036】

$$H(\omega) = (\text{スピーチのベクトル番号に対応したパワースペクトル}) / (\text{混合音声のパワースペクトル})$$
  
ウィナーフィルタ部507の出力スペクトルは、IFFT部508の高速逆フーリエ変換によって時間波形に復元される。この時間波形の周期的区間はくし形フィルタ部509へ、非周期的区間はつなぎ合わせ部510へ出力される。

【0037】くし形フィルタ部509では、IFFT部508の出力に対して、入力されたピッチデータに基づき、ピッチ周波数とその整数倍の周波数成分のみからな

るスピーチを取り出す。

【0038】上述のようにして復元された無音区間、周期的区間、非周期的区間の各信号波形は、つなぎ合わせ部510にて加算されて、分離されたスピーチとなる。

【0039】適応フィルタ部511では、分離されたスピーチに適応フィルタを施す。フィルタ出力を減算部512により混合音声から減算して背景音が分離される。分離された背景音は適応フィルタ部511にフィードバックされ適応フィルタ部511の出力が波形等価されるように最適な係数に修正される。

【0040】こうして混合音声から分離されたスピーチと背景音は前記混合器6へ出力され、この混合器6により聴取者の好みの割合で混合され最適な音声となるのである。

【0041】このように本実施例によれば、従来のスピーチ・エンハンスメントや雑音除去技術では困難であった受信側での信号分離が、信号分離に適した付加情報を送信側から伝送することによって可能となる。

【0042】また、従来技術では、混合音声から背景音を分離することは困難であったが、本実施例によれば、スピーチと背景音の両方を抽出分離でき、分離されたスピーチと背景音を聴取者の最も聞きやすい割合で混合することができる。したがって、放送システムに適用した場合、ニュース番組やドラマなど、多様な背景音を持つ番組を聴取するのに十分有効なものとなる。

【0043】

【発明の効果】以上説明したように本発明によれば、付加情報の容量を少なく抑えつつ、混合音声からスピーチと背景音とを高品質で分離でき、聴取者が自分の好みの割合で両者を混合することが可能となる。

【図面の簡単な説明】

【図1】本発明に係る混合音声信号伝送方式に全体構成を示すブロック図である。

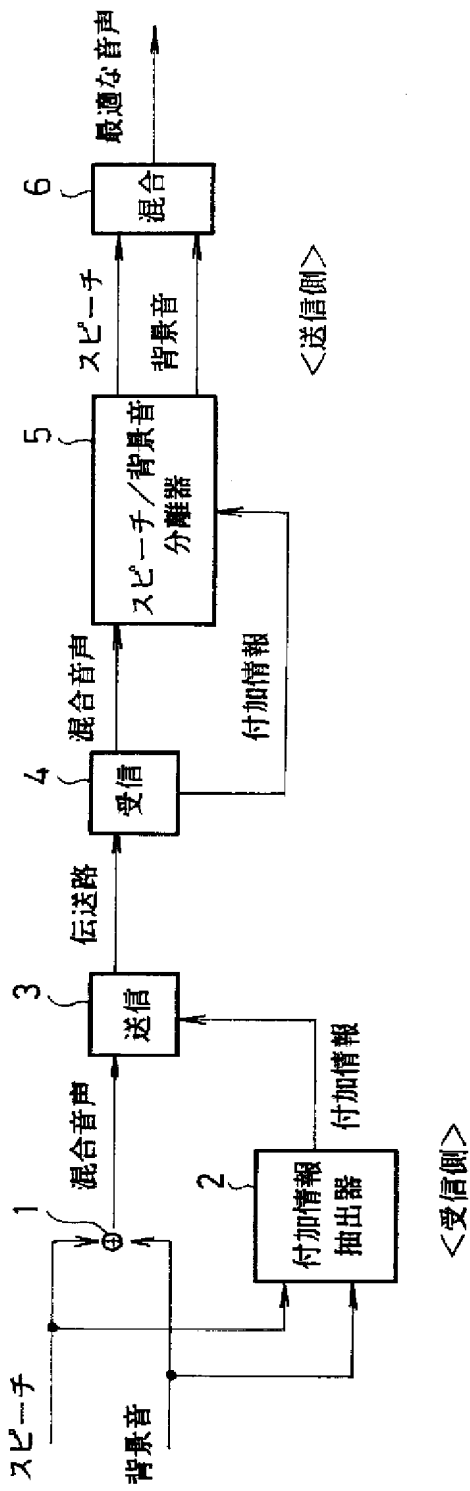
【図2】付加情報抽出器の一実施例構成を示すブロック図である。

【図3】スピーチ／背景音分離器の一実施例構成を示すブロック図である。

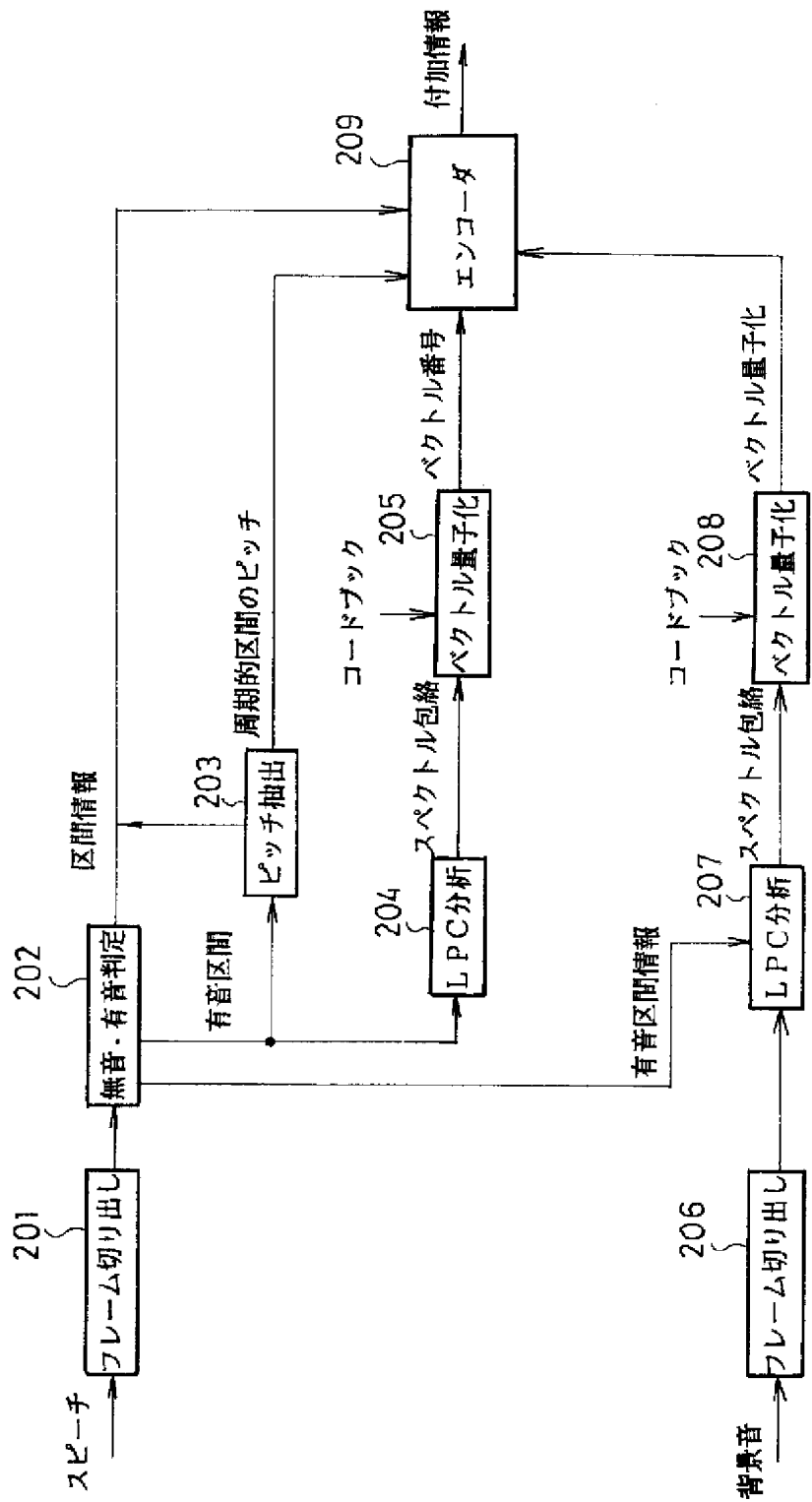
【符号の説明】

- 2 付加情報抽出器
- 3 送信機
- 4 受信機
- 5 スピーチ／背景音分離器
- 6 混合器

【図1】



【図2】



【図3】

